Probabilistic typhoon rainfall forecasting using a modified fuzzy inference model

Shien-Tsung Chen [*]

Department of Hydraulic and Ocean Engineering, National Cheng Kung University, Tainan City, Taiwan

ABSTRACT

Machine learning can yield timely forecasts of rainfall and flood hazard variables. Such an application of machine learning in disaster mitigation is a key topic in hydroinformatics. In this study, a modified fuzzy inference model—featuring revised implication and defuzzification processes—was used to perform probabilistic typhoon rainfall forecasting. The revised implication process exerted a higher weight of the rainfall variable than other typhoon variables on forecasting. The defuzzification process was modified into a resampling process to generate a predicted probability distribution. The modified fuzzy inference model was applied to real-time probabilistic typhoon rainfall forecasting with lead times of 1 to 3 hours. Predicted confidence intervals with respect to actual typhoon rainfall data validated the capability of the probabilistic forecasting method, although the predicted confidence intervals are wider than the perfect interval in view of a quantitative measure of the reliability diagram. In addition, the probabilistic forecasts could be condensed into deterministic forecasts using the median of the predicted confidence interval. The deterministic forecasts also had satisfactory performance in typhoon rainfall forecasting.

1. INTRODUCTION

Machine learning methods are extensively applied to hydrologic forecasting in hydroinformatics. Hydrologic forecasting provides timely hazard information to mitigate the impacts of natural disasters. Various machine learning methods have been applied to real-time hydrologic forecasting to forecast variables related to rainfall and flood. For example, artificial neural networks have been used to forecast flood (Dawson and Wilby 1998; Chang and Chen 2001; Chang et al. 2007; Jhong et al. 2018), adjust radar rainfall (Teschl et al. 2007; Chen et al. 2011), and predict reservoir inflow (Lin and Wu 2011; Sattari et al. 2012; Moeeni and Bonakdari 2017). Various methods based on fuzzy set theory have been applied to rainfall forecasting (Yu et al. 2004, 2005; Asklany et al. 2011) and flood forecasting (Yu and Chen 2005; Chen et al. 2013a, 2019). Grey system theory has been adopted to develop forecasting models that forecast rainfall (Yu et al. 2000), runoff (Yu et al. 2001), and flood stages (Chen 2015a). Support vector machines, popular in recent hydroinformatics research, have been widely used to develop real-time forecasting models that forecast flood stage and discharge (Bray and Han 2004; Yu et al. 2006; Chen and Yu 2007b; Han et al. 2007; Lin et al. 2013b). Support vector machines have also been applied to downscale hydrologic variables (Kim et al. 2018) and to project daily rainfall under climate change scenarios (Chen et al. 2010; Yang et al. 2011). In addition, a support vector machine was adopted as a data mining scheme to extract informative rainfall and flood data during flash floods (Chen 2015b). Recently, the self-organizing map and the long short-term memory has been successfully applied in hydrologic forecasting. For example, Chang et al. (2020) applied self-organizing maps of projected typhoon tracks to predict flood hydrographs prior to typhoon landfall. Chang et al. (2021) trained various self-organizing maps for clustering high-dimensional flood inundation maps. Kao et al. (2020, 2021) applied long short-term memory for multi-step-ahead flood and inundation forecasting.

The aforementioned studies demonstrate the prolific applications of machine learning methods in a variety of hydrologic forecasts. However, forecasting extreme cases such as typhoon rainfall is still a challenge in hydrologic

* Corresponding author
  E-mail: chen@gs.ncku.edu.tw

forecasting. Typhoons that accompany torrential rainfall often cause severe disasters such as landslide, debris flow, and floods. Typhoon rainfall forecasting that indicates future rainfall potentials is an important countermeasure in the mitigation of typhoon disasters caused by heavy rainfall. The method for forecasting typhoon rainfall can be categorized into the physical and statistical approaches. The physical approach uses the numerical weather prediction model developed on the basis of physical principles to forecast typhoon rainfall (e.g., Wu and Lin 2017; Luitel et al. 2018; Moses and Ramotonto 2018; Wu et al. 2018; Hsiao et al. 2020). The statistical approach includes the application of machine learning methods to forecast typhoon rainfall. Lin and Wu (2009) proposed a hybrid neural network model—combining a self-organizing map with a multilayer perceptron network—to forecast typhoon rainfall. Lin et al. (2013a) adopted support vector machines with a multiobjective genetic algorithm to develop a typhoon rainfall forecasting model for forecasting hourly typhoon rainfall in real time. Chen (2013) applied multiclass support vector classification to estimate the temporal rainfall distribution of a typhoon upon the issuance of a typhoon warning. Although research has been promising, an advance in real-time typhoon rainfall forecasting is the development of probabilistic forecasting methods for forecasting typhoon rainfall.

Krzysztofowicz (2001) discussed the advantages of probabilistic forecasting over deterministic forecasting in hydrology. In addition to providing economic benefits, probabilistic forecasts possess greater veracity, enable risk-based warnings, and aid decision-making under uncertainty. Bliefernicht and Bárdossy (2007) used statistical downscaling techniques to produce probabilistic daily precipitations focusing on extreme events. Chen and Yu (2007a) proposed a probabilistic forecasting method comprising a deterministic forecast derived from support vector regression and a probability distribution of forecast error based on the fuzzy inference model. They applied the probabilistic forecasting method to real-time flash flood forecasting and attained practical probabilistic forecasting results. Villarini et al. (2010) used ensemble techniques that combine individual output values forecasted by artificial neural networks, and they applied the method to peak flood discharge forecasting. Chen (2019) used the probabilistic forecasting method and the modified fuzzy inference model to produce probabilistic forecasts of coastal wave heights during a typhoon warning period. Chen et al. (2013b) applied a Markov chain process to develop a probabilistic drought forecasting model based on the teleconnection of drought indices and sea surface temperatures. Araghinejad et al. (2006) developed a geostatistically based approach featuring a local regression method to probabilistically forecast streamflow using ocean-atmospheric signals and hydrologic conditions. Nguyen and Chen (2020) applied multiple machine learning methods to probabilistically forecast hourly river

level during the flash flood. Doong et al. (2020) used artificial neural network to develop an operational probabilistic forecasting system to predict the probability of coastal freak wave occurrence. Zhou et al. (2020) applied the unscented Kalman filter with recurrent neural networks to improve the reliability of probabilistic flood forecasting. In the field of hydrologic forecasting, although some applications in probabilistic forecasting have been proposed, this study is the first to use machine learning methods in the probabilistic forecasting of typhoon rainfall.

Among the various machine learning methods, the fuzzy inference model was adopted in this study. Although the fuzzy inference process can be criticized for being imprecise, its rule-based logic is simple and understandable. The fuzzy inference model can easily translate an observed dataset into a fuzzy rule. This is particularly suitable for constructing a database of many fuzzy rules for probabilistic forecasting. To better generate probabilistic forecasts of typhoon rainfall, this study applied a modified fuzzy inference model. The modifications of the fuzzy inference model included the calculation of resulting membership grade in the implication process and the transformation of the defuzzification process into a probability distribution. Section 2 of this paper provides the methodology of the modified fuzzy inference model. Section 3 describes the study area and hourly rainfall and typhoon data. Section 4.1 presents the development of probabilistic typhoon rainfall forecasting model, and section 4.2 describes the probabilistic forecasting results pertaining to validation events. The validation results confirm the ability of the modified fuzzy inference model. Moreover, deterministic forecasts that can be easily obtained from the median of the predicted confidence interval (CI) were proposed and demonstrated in section 4.3. Section 5 concludes the paper in that both probabilistic and deterministic forecasts exhibited satisfactory forecasting performance.

2. MODIFIED FUZZY INFERENCE MODEL

A fuzzy inference model is a system that uses fuzzy logic to emulate how a human expert reasons. The original version of the modified fuzzy inference model has been proposed in Chen and Yu (2007a) and Chen (2019). The modification centers on changing the defuzzification procedure into a resampling process that can generate a probability distribution. In this study, the implication process uses a hybrid operation to calculate the resulting membership grade. The methodology of the modified fuzzy inference model is as follows.

2.1 Defining the Membership Function and Fuzzifying Variables

The first step in performing fuzzy inference is defining

the fuzzy membership function. A fuzzy membership function can determine the level of membership of an element in a set to a target. The level of membership, which is termed membership grade, is defined within the unit interval [0, 1]. An example of triangular membership function can be found in Fig. 1. A membership grade of 0 means that an element does not belong to that target. A membership grade of 0.8, for example, indicates that an element belongs to that target to a high degree. The Gaussian membership function, widely used to fuzzify the variables, was adopted. Specifically,

$$\mu(x) = \exp\left[\frac{-(x-m)^2}{2\sigma^2}\right] \tag{1}$$

where the parameter σ is the extent of the membership function; $(x - m)$ is the difference between an input value x and the center of the membership function m; and $\mu(x)$ is the membership grade. In this study, the parameter σ is the only parameter to be calibrated. The degree of similarity between x and m can be represented by a membership grade that ranges from 0 to 1. When x is close to m, the degree of similarity is large, and the membership grade is close to 1; when x is not close to m, the membership grade and similarity decay exponentially to zero.

2.2 Formulating the Fuzzy Rule and Constructing the Fuzzy Rule Database

A fuzzy rule is formulated as an "IF-THEN" statement, where the "IF" component is the antecedent (or premise) of the fuzzy rule, and the "THEN" component is the consequence (or conclusion) of the rule. A fuzzy rule can be expressed as

$$\begin{aligned} &\text{IF } (x_1 \text{ is } A_1) \text{ AND } (x_2 \text{ is } A_2)\, ... \text{ AND } (x_m \text{ is } A_m)\\ &\text{THEN } (y_1 \text{ is } B_1) \text{ AND } (y_2 \text{ is } B_2)\, ... \text{ AND } (y_n \text{ is } B_n) \end{aligned} \tag{2}$$

where $x_1, x_2, \ldots, x_m$ are input variables; $y_1, y_2, \ldots, y_n$ are output variables; and $A_1, A_2, \ldots, A_m$ and $B_1, B_2, \ldots, B_n$ denote fuzzy sets that are defined by membership functions.

If the antecedent part of the rule is true to a certain degree, which is the membership grade defined by the membership function, then the consequence part is true to that same degree. The fuzzy rule can be formulated as either a linguistic statement or a direct numerical relation between input and output data. This study directly used environmental observations to formulate fuzzy rules. That is, each collected dataset was used to create the direct numerical relation in the form of a fuzzy rule. These generated fuzzy rules, which were assembled and stored in a fuzzy rule database, characterize the fuzzified relationship between the input and output data.

2.3 Calculating Similarity by Fuzzy Implication

The fuzzy implication process calculates the similarity between input data and the fuzzy rules (i.e., the fuzzified observations stored in the database). For each fuzzy rule, the implication process generates a truncated fuzzy subset (shown in Fig. 1 as the blue trapezoid for example) of the output variable in the consequence part according to a resulting membership grade in the antecedent part. The resulting membership grade essentially depicts the similarity between input data and observations in the database. The left part of Fig. 1 illustrates the fuzzy implication process with two input variables in the antecedent part and one output variable in the consequence part of the fuzzy rule. In Fig. 1, two membership grades of two input variables are obtained according to the membership functions in the fuzzy rule. Typically, a fuzzy intersection operation is used to select the minimum membership grade for the output variable. However, this study applied the average operation to obtain the average membership grade to truncate the output fuzzy set. The average membership grade is a more appropriate measure of similarity than the minimum membership grade. This is because the selection of the minimum membership grade is necessarily dominated by the least similarity value.

2.4 Aggregating Similarity Values to Create a Similarity Database

Aggregation is a process for integrating the truncated fuzzy subset of each fuzzy rule from the implication procedure into a single aggregated fuzzy set. Aggregation is conducted using the fuzzy union operation (see the middle part in Fig. 1). The input of the aggregation process is a group of truncated fuzzy sets (for example, the two trapezoids in Fig. 1), and the output is a single, combined fuzzy set. The aggregation procedure in the modified fuzzy inference model represents the creation of a similarity database through the accumulation of the similarity values of each fuzzy rule. In this database, the following resampling procedure is performed to sample out the favored data.

2.5 Resampling to Construct a Probability Distribution

After creating the similarity database, an input dataset can be made to correspond to each fuzzy rule (fuzzified observations) with a similarity value. When an input dataset is considered, the model output is more and less likely to be the output variable in a fuzzy rule of high and low similarity, respectively. To cover the possibility of all outcomes (in order to form a probability distribution), it is rational to use a resampling scheme to obtain the possible output data from the similarity database (see the right part in Fig. 1). Naturally, the output variable in a fuzzy rule of higher similarity has a higher probability to be sampled out relative to
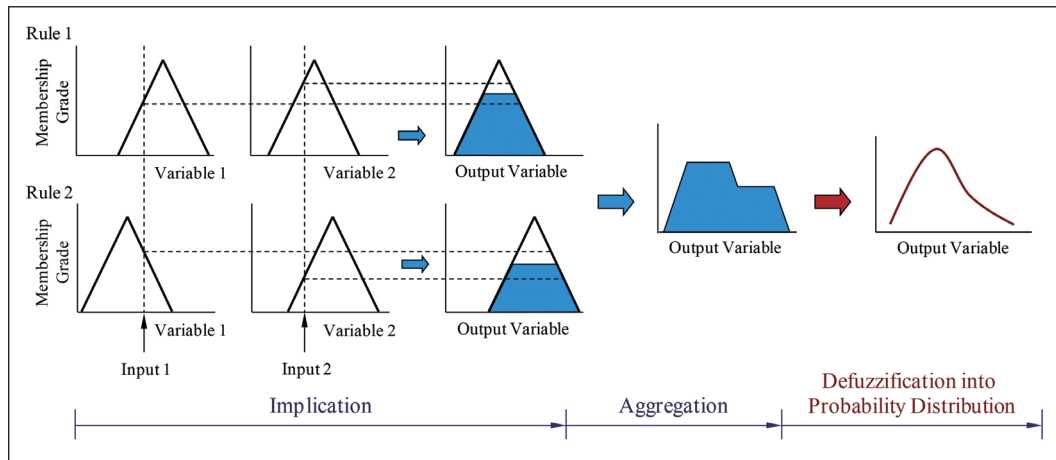
Fig. 1. Modified fuzzy inference model.

that of lower similarity. The probability to be sampled can be determined with respect to the similarity values and is expressed as

$$p_i = \frac{\mu_i}{\sum_{i=1}^{l} \mu_i} \qquad (3)$$

where p_i denotes the probability to be sampled, μ_i is the similarity value in line with the i-th fuzzy rule, and l is the number of fuzzy rules. Each dataset (fuzzy rules) in the similarity database has its own probability p_i to be sampled out. These data and their corresponding probabilities can form a possibility distribution. The resampling scheme is, in concept, the same as Monte Carlo simulation. By using a repeated resampling process, the desired number of the values regarding the output variable can be obtained, and the probability distribution of the output variable can be constructed. Subsequently, the confidence interval (CI) with different confidence levels can be defined from the constructed probability distribution.

3. STUDY AREA AND DATA

Taiwan is an island in the northwest Pacific. The Tropic of Cancer lies across the middle of Taiwan, making the climate tropical in the south and subtropical in the north. With its particular geographic location, Taiwan is often affected by typhoons—on average, four yearly. The typhoon that generates on the northwest Pacific Ocean often moves directly toward the east coast of Taiwan, resulting in heavy rainfall. Hualien, the biggest city in the east coast, was chosen as the study area (Fig. 2). Long-term hourly precipitation data were available from the Hualien Meteorological Station operated by the Central Weather Bureau (CWB) of Taiwan.

When a typhoon approaches Taiwan, the CWB of Tai-

wan issues typhoon warnings, and related meteorological data will be archived in the CWB typhoon database. Typhoon data such as central pressure, maximum sustained wind speed, storm radius (with wind speed of 17.2 m s^{-1}), and latitude and longitude are available from the CWB typhoon database. Formally, the CWB issues typhoon warnings every three hours. But in practice, the typhoon warning was issued every hour when a typhoon substantially influences Taiwan. In circumstances when hourly data are not available from CWB, a linear interpolation scheme was used in this study to obtain hourly typhoon data. Therefore, this study compiled hourly rainfall and typhoon data during the typhoon warning period issued by CWB. Table 1 lists the collected events from 2006 to 2019 and their rainfall and typhoon characteristics during the typhoon warning period, including the date, name of the typhoon, number of data points, minimum central pressure, maximum storm radius, peak intensity of the maximum sustained wind speed, moving tracks, and total rainfall depth. Figure 3 displays the 10 types of typhoon tracks categorized by the CWB. The typhoon rainfall amount of collected events during the warning periods ranged from 30.5 to 513.6 mm. Hourly data of typhoon center position (longitude and latitude) were also compiled. Therefore, the position of a typhoon relative to Hualien Meteorological Station (121.61°E, 23.98°N) can be determined using a polar coordinate system, with the origin set at Hualien. The position of typhoon is thus identified by two variables: distance and angle. The distance is the great-circle distance on the surface of the Earth. The angle facing the east is 0°, with positive values in the counterclockwise direction and negative values in the clockwise direction. In total, 45 typhoon events with complete rainfall and typhoon data were identified. The first 36 events were used as calibration events to construct the fuzzy rule database, and the remaining nine events were used as validation events to examine the proposed modified fuzzy inference model.

4. RESULTS AND DISCUSSION

4.1 Model Development

This study applied the modified fuzzy inference model to perform typhoon rainfall forecasting. Typhoon data used as model input variables are the central pressure (P), maximum sustained wind speed (V), storm radius (S), great-circle distance (D), and angle (A). Although the rainfall is not very sensitive to the pressure and radius, including more typhoon information in the fuzzy inference model is preferred. The rainfall variable used in the model is the cumulative rainfall (R). That is, the predictand in this study is the accumulated rainfall from the beginning of the first typhoon warning issued by CWB. The cumulative rainfall (R) at the preceding time is also used as an input to enhance forecasting performance. At each forecasting hour, the forecasting lead times were 1 to 3 hours. Namely, the proposed typhoon rainfall forecasting model generates future 1- to 3-hour probabilistic forecasts of cumulative rainfall at every hour during the typhoon warning period. The architecture of the probabilistic typhoon rainfall forecasting model can be formulated in terms of a fuzzy rule as follows.

$$IF\,[P(t)\text{ is }F_P]\text{ AND }[V(t)\text{ is }F_V]\text{ AND }[S(t)\text{ is }F_S]$$
$$\quad AND\,[D(t)\text{ is }F_D]\text{ AND }[A(t)\text{ is }F_A]\text{ AND }[R(t)\text{ is }F_R]$$
$$THEN\,[\widehat{R}(t+1)\text{ is }F_{R(t+1)}]\text{ AND }[\widehat{R}(t+2)\text{ is }F_{R(t+2)}]\quad(4)$$
$$\quad AND\,[\widehat{R}(t+3)\text{ is }F_{R(t+3)}]$$

where F_P, F_V, F_S, F_D, F_A, and F_R, respectively, represent the defined fuzzy sets of the variables of pressure (P), speed (V), radius (S), distance (D), angle (A), and rainfall (R) at time t. $\widehat{R}(t+1)$, $\widehat{R}(t+2)$, and $\widehat{R}(t+3)$ denote the 1- to 3-hour probabilistic forecasts generated by the modified fuzzy inference model. Alternatively, the typhoon rainfall forecasting model can be formulated in terms of functions as follows.

$$\widehat{R}(t+1)=f_1[P(t),V(t),S(t),D(t),A(t),R(t)]\quad(5)$$

$$\widehat{R}(t+2)=f_2[P(t),V(t),S(t),D(t),A(t),R(t)]\quad(6)$$

$$\widehat{R}(t+3)=f_3[P(t),V(t),S(t),D(t),A(t),R(t)]\quad(7)$$

where f_1, f_2, and f_3 are the mechanisms of the modified fuzzy inference model that infer the probabilistic forecasts of lead times of 1, 2, and 3 hours, respectively.

The defuzzification of variables requires determining the parameter of the membership function in Eq. (1). Because the parameter σ of the Gaussian function represents the spread of the membership function, the standard deviation of the calibration data has been recommended as the

parameter value (Chen and Yu 2007a). The standard deviations of the pressure (P), speed (V), radius (S), distance (D), angle (A), and rainfall (R) are, respectively, 23.2 hPa, 10.4 m s^{-1}, 59.8 km, 191.9 km, 99.3°, and 111.3 mm. This study directly used standard deviations as parameter values, except for the rainfall variable. The rainfall variable used in the model is the cumulative rainfall, where the values are very small in the beginning and can be very large (in the hundreds) at the end of an event. Specifically, the standard deviation of rainfall variable is 111.3 mm, which is a very large value for rainfall. Using such a large parameter in the membership function makes membership grades too big for most rainfall data. Therefore, this study applied a simple grid search method to optimize the parameter of the membership function for rainfall to be 20 mm.

The implication process of the modified fuzzy inference model in this study adopted the average operation to obtain the resulting membership grade (similarity). Experience from the analytical process revealed that to forecast future rainfall, the same variable (rainfall) should be weighted higher than other input variables to attain better forecasting performance. Therefore, a hybrid operation is used to calculate the resulting membership grade μ. Specifically, first, a membership grade μ_1 is determined as the minimum membership grade for the variables of pressure (P), speed (V), radius (S), distance (D), and angle (A) and denoted μ_P, μ_V, μ_S, μ_D, and μ_A, respectively. Subsequently, a membership grade μ_2 is calculated for rainfall (R). The averaging of μ_1 and μ_2 yields the resulting membership grade μ. Specifically,

$$\mu_1=\min[\mu_P,\mu_V,\mu_S,\mu_D,\mu_A]\quad(8)$$

$$\mu=\text{average}[\mu_1,\mu_2]\quad(9)$$

When the input and output variables in Eq. (4) and their corresponding parameters are determined, calibration data can be used to construct the fuzzy database. Then, the probabilistic typhoon rainfall forecasting was performed with respect to validation events. At each time t during the typhoon warning period, future probabilistic forecasts with lead times of 1 to 3 hours could be obtained by the modified fuzzy inference model. The resampling procedure was applied to sample out 10000 rainfall forecasts from the fuzzy database. The Weibull plotting position formula was used to make these 10000 forecasts into the predictive probability distribution, from which different percentages of the CI can be directly derived. The probabilistic typhoon rainfall forecasting results are as follows.

4.2 Probabilistic Typhoon Rainfall Forecasting Results

Figure 4 illustrates an example of probabilistic forecasting with different CIs. The solid line represents the

observed cumulative rainfall during the typhoon warning period. The shaded area represents the CI. The blue shaded area represents the forecasted 90% CI (i.e., out of the 10000 forecasts, 9000 forecasts lie within the upper and lower bounds of the interval, and 1000 forecasts are outside the blue shaded area). Thus, the predicted 99% CI has a wider range than the predicted 90% CI. Because the 90% CI can well enclose the observations, good forecasting performance is indicated; the model yields a high confidence that the observation will appear within the predicted interval.

Figures 5 to 7 illustrate the probabilistic forecasts with lead times of 1 to 3 hours, respectively. The predicted 90% CI widens as lead time increases, indicating that 3-hour forecasting has greater uncertainty relative to 1-hour forecasting. The 90% CI well encloses the observations, except for some cases. An instance of a 3-hour forecast at the end of Event 38 is discussed. At the 55th hour of Event 38, the observed cumulative rainfall is 174 mm. A sudden increase from heavy rainfall yields future rainfall observations of 202, 220, and 241 mm for the following 3 hours. The rainfall increments are as large as 28, 46, and 67 mm. Figure 8 illustrates the 1- to 3-hour forecasted probability distribution at the 55th hour of Event 38. For 3-hour forecasting, the 90% CI is [139.0, 226.9], not including the future observation of 241 mm. However, the 95% CI [139.0, 247.1] can enclose the observation. The observation of 241 mm corresponds to the forecasted probability distribution at a cumulative probability value of 0.973. That is, the proposed model gives a 2.7% chance that future cumulative rainfall will exceed 241 mm after 3 hours. Although the forecasting is not exceedingly accurate for this extreme case, the probabilistic forecasting results remind users of the uncertainty in extreme cases, thus allowing the decision maker to implement risk-based countermeasures.

An important measure for assessing the performance of probabilistic forecasting is the accurate enclosure of the amount of data within the CI. If the predictive probability distribution can effectively explain the total uncertainty, then the percentage of data included in the CI is equivalent to the probability (confidence level) of the CI. Figure 9 shows the reliability diagram of the probabilistic forecasting results for all validation events. In Fig. 9, the forecasted confidence interval indicates the confidence level of the CI, and the observed frequency signifies the percentage of observed data included in the CI. A 45-degree line in the reliability diagram means perfect forecasts in view of probabilistic forecasting. According to Fig. 9, a larger amount of data than expected are enclosed in the CI. For example, in 1-hour forecasting, a forecasted CI with confidence level of 60% (forecasted probability of 60%) encloses 90% of the observations (observed probability of 90%). This indicates that the predictive CIs are wider than perfect intervals. This phenomenon can also be seen from Figs. 5 to 7 in that wide CIs are forecasted for zero or small rainfalls in the beginning of an event and for small events. Moreover, the performance of 2- and 3-hour probabilistic forecasting is inferior to that of 1-hour forecasting, because the observed probabilities for 2- and 3-hour forecasting are even larger than that for 1-hour forecasting in Fig. 9. Although the proposed probabilistic forecasting model is not perfect in view of the reliability diagram, the median of the CI well corresponds to the observed rainfall, which is described in the following section.

Fig. 2. Location of study area.

Table 1. Typhoon events and rainfall data.

Event No.	Date (yyyy/mm/dd)	Name of Typhoon	No. of Data	Minimum Pressure (hPa)	Maximum Storm Radius (km)	Peak Intensity (m s⁻¹)	Moving Track	Total Rainfall (mm)
1	2006/05/16	Chanchu	48	943	300	45	9	154.5
2	2006/07/12	Bilis	73	978	300	25	2	42.5
3	2006/07/23	Kaemi	37	935	180	48	3	74.5
4	2006/08/07	Bopha	39	988	120	23	4	196.5
5	2007/08/06	Pabuk	36	980	150	28	4	393.5
6	2007/08/08	Wutip	25	992	100	18	3	53.5
7	2007/08/16	Sepat	79	920	250	53	3	351.9
8	2007/10/04	Krosa	79	925	300	51	2	79.7
9	2007/11/26	Mitag	30	965	200	35	*	205.5
10	2008/07/16	Kalmaegi	58	970	120	33	2	212.0
11	2008/07/26	Fung-Wong	73	948	220	43	3	405.5
12	2008/09/11	Sinlaku	127	925	250	51	2	139.0
13	2008/09/21	Hagupit	43	940	280	45	*	40.0
14	2008/09/26	Jangmi	73	925	280	53	2	254.4
15	2009/06/19	Linfa	122	980	150	28	9	74.5
16	2009/07/16	Molave	73	980	100	28	*	64.5
17	2009/08/05	Morakot	208	955	250	40	3	155.6
18	2009/10/03	Parma	171	945	250	43	10	513.6
19	2010/08/31	Lionrock	43	990	100	23	9	71.5
20	2010/09/09	Meranti	26	990	100	23	*	97.5
21	2010/09/17	Fanapi	64	940	200	45	4	242.1
22	2010/10/21	Megi	70	935	250	48	9	87.7
23	2011/05/27	Songda	37	920	220	55	*	60.0
24	2011/08/27	Nanmadol	100	920	180	53	4	358.0
25	2012/06/19	Talim	49	985	150	25	9	30.5
26	2012/06/28	Doksuri	28	995	120	23	*	42.0
27	2012/07/30	Saola	91	960	220	38	2	407.5
28	2012/08/21	Tembin	97	945	180	45	10	144.5
29	2013/08/20	Trami	46	970	180	30	1	34.6
30	2013/09/19	Usagi	63	910	280	55	5	195.0
31	2013/10/04	Fitow	58	960	250	38	1	50.0
32	2014/07/21	Matmo	55	960	200	38	3	324.5
33	2014/09/19	Fung-Wong	73	985	150	25	10	156.0
34	2015/08/06	Soudelor	70	930	300	48	3	199.0
35	2015/08/20	Goni	76	910	220	51	*	233.0
36	2015/09/27	Dujuan	58	925	200	51	2	181.5
37	2016/07/06	Nepartak	72	905	200	58	4	281.0
38	2016/09/12	Meranti	61	900	220	60	7	269.5
39	2016/09/25	Megi	67	940	250	45	3	284.0
40	2017/07/28	Nesat	51	955	180	40	2	94.0
41	2017/07/30	Haitang	22	990	100	20	7	44.5
42	2017/08/20	Hato	43	965	180	33	*	87.0
43	2018/09/14	Mangkhut	34	895	320	60	*	73.0
44	2019/08/23	Bailu	55	975	150	30	4	209.5
45	2019/09/29	Mitag	52	960	180	38	6	51.5

*Note: Asterisk * indicates that the typhoon did not make landfall and its track was not classified.*
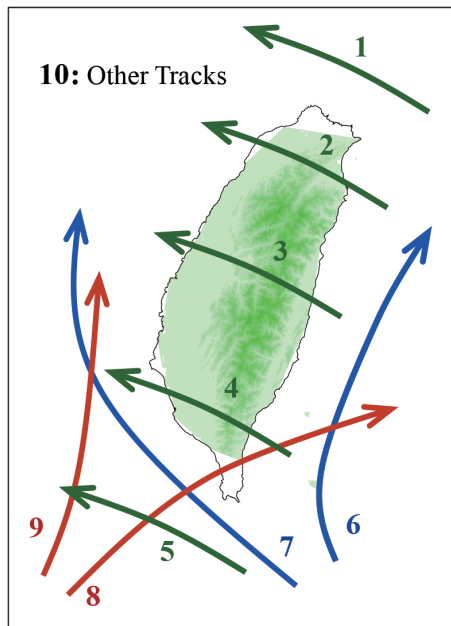
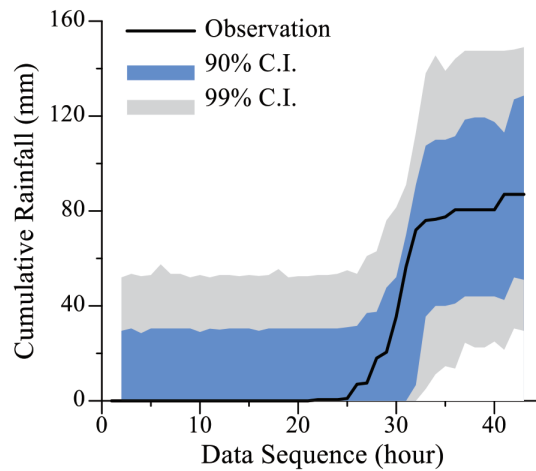Fig. 3. Moving tracks of typhoons classified by CWB.

Fig. 4. Probabilistic typhoon rainfall forecasting for Event 42 with 90 and 99% confidence intervals (CIs).
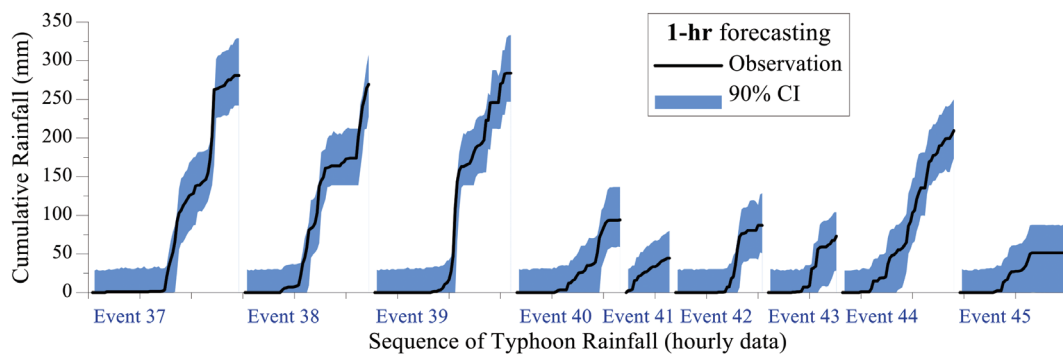

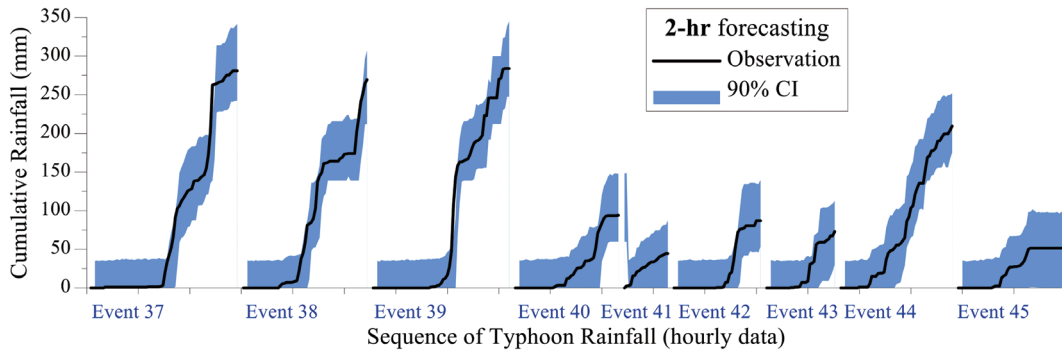Fig. 5. Probabilistic typhoon rainfall forecasting with lead time of 1 hour.

Fig. 6. Probabilistic typhoon rainfall forecasting with lead time of 2 hours.
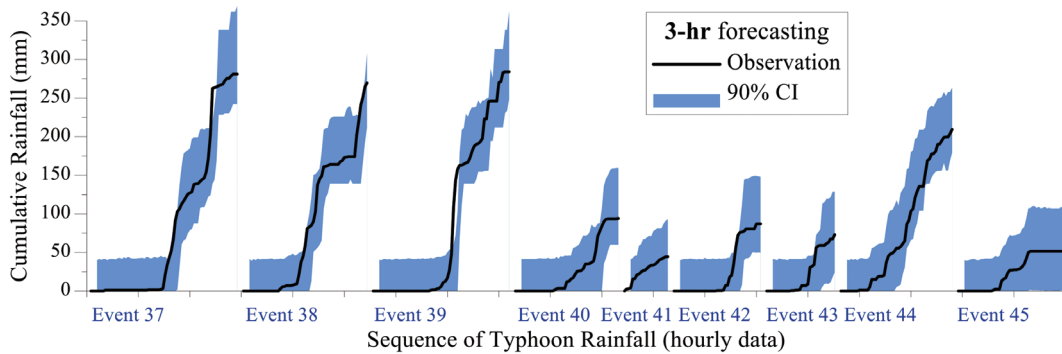
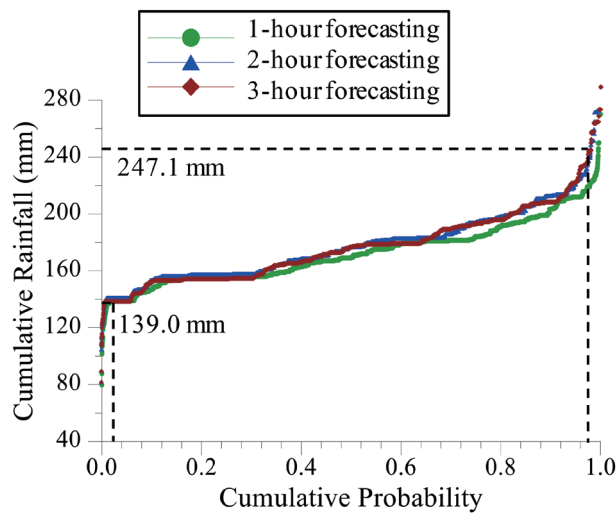Fig. 7. Probabilistic typhoon rainfall forecasting with lead time of 3 hours.

Fig. 8. Example of forecasted probability distribution with lead times of 1 to 3 hours, and 95% confidence interval for 3-hour forecasting.
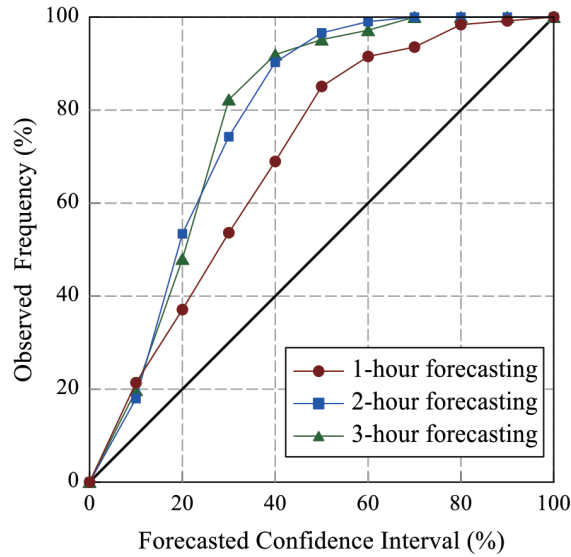
Fig. 9. Reliability diagram of the probabilistic forecasting results.

4.3 Deterministic Forecasting as a Supplementary Forecasting Product

Unlike deterministic forecasting, the nature of probabilistic forecasting makes the quantitative assessment of forecasting performance difficult. However, the proposed probabilistic forecasting method generates numerous possible forecasts to establish a predicted probability distribution, allowing it to be easily transformed into a deterministic forecasting method. A simple and rational scheme for this transformation involves picking the median out of the generated forecasts to form the deterministic forecast. Figure 10 shows the deterministic forecasting of the cumulative rainfall of Event 38. The deterministic forecasts lay close to the middle of the predicted CIs, with reference to the shaded area in Figs. 5 to 7. The deterministic forecasts resembled the observations with a minor phase lag, especially in 3-hour forecasting. The phase lag reveals that the rainfall forecasts are dominated by the preceding rainfalls. This is reasonably due to enhancing the weight of rainfall variable in the fuzzy inference process.

Statistical indices can be easily applied to deterministic forecasting to assess forecasting performance. This study used the correlation coefficient (CC) and the mean percentage error (MPE) to assess the performance of deterministic cumulative rainfall forecasts. The CC is a commonly used index for assessing the correlation between observations and forecasts. The MPE that calculates the relative error is suitable for evaluating performance with regard to cumulative rainfall forecasts. The CC and MPE are formulated as follows:

$$CC = \frac{\sum_{i=1}^{n}(R_i - m)(\widetilde{R}_i - \widetilde{m})}{\sqrt{\sum_{i=1}^{n}(R_i - m)^2}\sqrt{\sum_{i=1}^{n}(\widetilde{R}_i - \widetilde{m})^2}} \quad (10)$$

$$MPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{R_i - \widetilde{R}_i}{R_i} \quad (11)$$

where R is the observed cumulative rainfall, m is the mean of R, $\widetilde{R}$ is the deterministic forecast for cumulative rainfall, $\widetilde{m}$ is the mean of $\widetilde{R}$, and n is the number of data points. The cumulative rainfall variable, being an accumulated quantity, has very small and very large values. Thus, performance assessment in relation to cumulative rainfall is appropriate only if the relative error is used. Table 2 lists performance indices of the deterministic forecasting with respect to cumulative rainfall for all validation events. The CCs for 1- to 3-hour forecasting are above 0.98. The MPE values indicate that the deterministic cumulative rainfall forecasts had an approximately 30% error relative to observation. The calculation of MPE does not include observed values of cumulative rainfall less than 10 mm. When the observed rainfall is small and in the denominator, the MPE index can be biased.

In addition to cumulative rainfall, the rainfall amount at a time step is often of public and hydrologic interest. The hourly cumulative rainfall can be easily transformed into hourly rainfall by subtracting the preceding cumulative rainfall from the current cumulative rainfall. Figure 11 details the deterministic forecasts for the hourly rainfall of Event 38. Although hourly rainfall is somewhat random, the deterministic forecasts could detect the variation in hourly rainfall with a minor phase lag. To evaluate forecasting performance
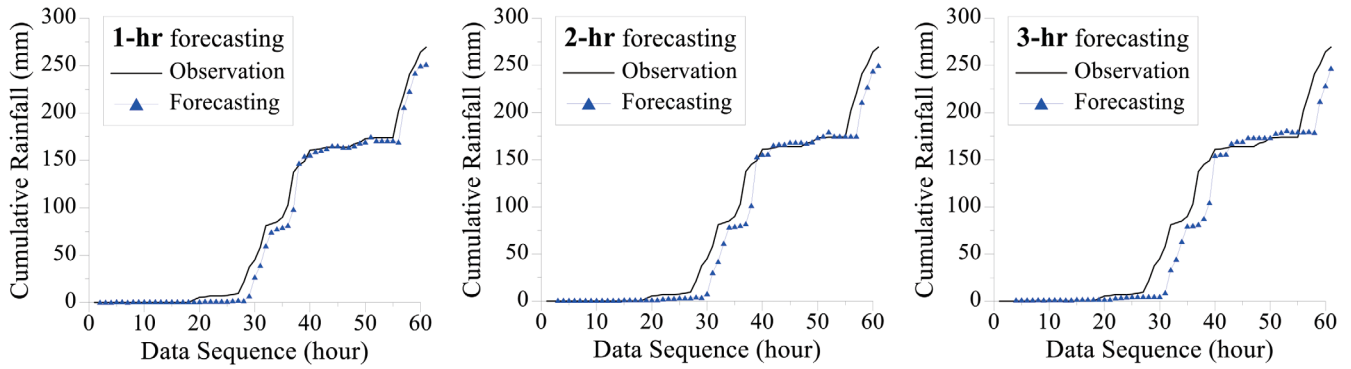
Fig. 10. Deterministic forecasting for cumulative rainfall for Event 38.

Table 2. Performance indices for deterministic fore-casting for cumulative rainfall.

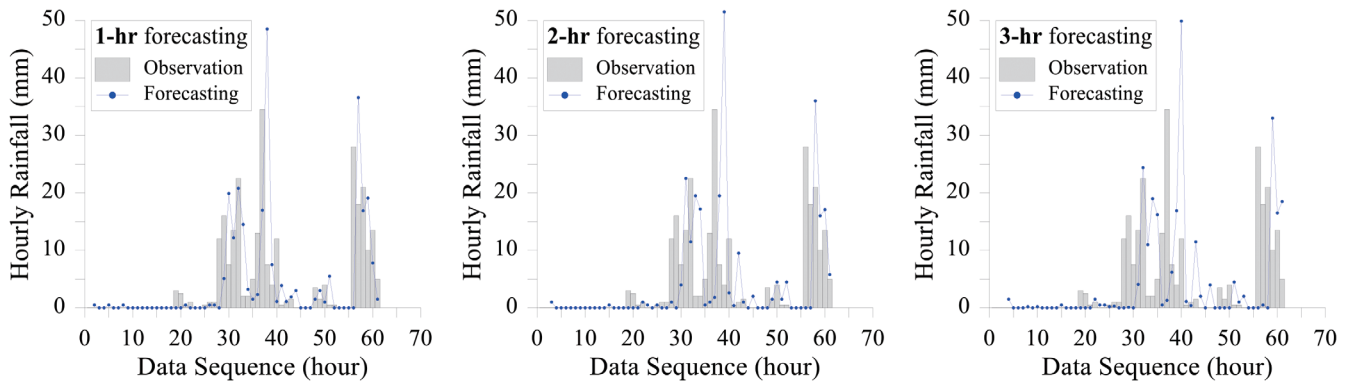|	CC	MPE (%)
1-hour forecasting	0.99	23.7
2-hour forecasting	0.98	24.6
3-hour forecasting	0.98	27.6

Fig. 11. Deterministic forecasting for hourly rainfall of Event 38.

in the case of hourly rainfall, CC and the mean absolute error (MAE) are calculated. The MAE is a direct measure of error between forecast and observation. Specifically,

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| r_i - \tilde{r}_i \right|}{n} \tag{12}$$

where r is the observed hourly rainfall, and $\tilde{r}$ is the deterministic forecast for hourly rainfall. Table 3 lists the CC and MPE with respect to hourly rainfall for all validation events. The CC is 0.57 for 1-hour forecasting and 0.19 for 3-hour forecasting. The CCs are not high because rainfall is more random in hourly than in cumulative rainfall. The MAEs range from 3.3 to 3.8 mm, indicating that the

error for hourly rainfall is not large. The inferior predictability of hourly rainfall can be related to the phase lag of the predicted cumulative rainfall. Reducing the weight of the rainfall variable during the fuzzy inference process may possibly diminish the phase lag and improve the deterministic forecasts. However, the main objective of this study is the probabilistic forecasting. Therefore, the emphasis is put on the performance of probabilistic forecasts. Overall, the deterministic forecasting method that can supplement forecasting information exhibited reasonable performance.

5. CONCLUSION

This study proposed a probabilistic typhoon rainfall forecasting method using a modified fuzzy inference model

Table 3. Performance indices for deterministic fore-
casting method for hourly rainfall.

	CC	MAE (mm)
1-hour forecasting	0.57	3.7
2-hour forecasting	0.35	3.3
3-hour forecasting	0.19	3.8

with rainfall and typhoon data as input. The implication process uses a hybrid operation to calculate the resulting membership grade. Specifically, the weight of the rainfall variable is increased to improve forecasting performance. Defuzzification in the modified fuzzy inference model is a resampling process that generates many forecasts to form a predicted probability distribution. Therefore, the predicted CIs for different percentages can be constructed as probabilistic forecasting products. Validation results using data on typhoon rainfall in Hualien, Taiwan—where typhoons often bring heavy rainfall—verify the capability of the model to predict probabilistic forecasts with lead times of 1 to 3 hours. Moreover, the probabilistic forecasts can be condensed into deterministic forecasts by taking the median out of the predicted CI. The deterministic forecasting results also indicated satisfactory forecasting performance. As a probabilistic forecasting method, the modified fuzzy inference model has great potential to be used in various scholarly fields and practical applications.

REFERENCES

Araghinejad, S., D. H. Burn, and M. Karamouz, 2006: Long-lead probabilistic forecasting of streamflow using ocean-atmospheric and hydrological predictors. *Water Resour. Res.*, **42**, W03431, doi: 10.1029/2004WR003853. [Link]

Asklany, S. A., K. Elhelow, I. K. Youssef, and M. Abd El-wahab, 2011: Rainfall events prediction using rule-based fuzzy inference system. *Atmos. Res.*, **101**, 228-236, doi: 10.1016/j.atmosres.2011.02.015. [Link]

Bray, M. and D. Han, 2004: Identification of support vector machines for runoff modelling. *J. Hydroinform.*, **6**, 265-280, doi: 10.2166/hydro.2004.0020. [Link]

Bliefernicht, J. and A. Bárdossy, 2007: Probabilistic forecast of daily areal precipitation focusing on extreme events. *Nat. Hazards Earth Syst. Sci.*, **7**, 263-269, doi: 10.5194/nhess-7-263-2007. [Link]

Chang, F.-J. and Y.-C. Chen, 2001: A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *J. Hydrol.*, **245**, 153-164, doi: 10.1016/S0022-1694(01)00350-X. [Link]

Chang, F.-J., Y.-M. Chiang, and L.-C. Chang, 2007: Multi-step-ahead neural networks for flood forecasting. *Hy-*

drol. Sci. J., **52**, 114-130, doi: 10.1623/hysj.52.1.114. [Link]

Chang, L.-C., F.-J. Chang, S.-N. Yang, F.-H. Tsai, T.-H. Chang, and E. E. Herricks, 2020: Self-organizing maps of typhoon tracks allow for flood forecasts up to two days in advance. *Nat. Commun.*, **11**, doi: 10.1038/s41467-020-15734-7. [Link]

Chang, L.-C., W.-H. Wang, and F.-J. Chang, 2021: Explore training self-organizing map methods for clustering high-dimensional flood inundation maps. *J. Hydrol.*, **595**, 125655, doi: 10.1016/j.jhydrol.2020.125655. [Link]

Chen, C.-S., Y.-D. Jhong, T.-Y. Wu, and S.-T. Chen, 2013a: Typhoon event-based evolutionary fuzzy inference model for flood stage forecasting. *J. Hydrol.*, **490**, 134-143, doi: 10.1016/j.jhydrol.2013.03.033. [Link]

Chen, C.-S., Y.-D. Jhong, W.-Z. Wu, and S.-T. Chen, 2019: Fuzzy time series for real-time flood forecasting. *Stoch. Environ. Res. Risk Assess.*, **33**, 645-656, doi: 10.1007/s00477-019-01652-8. [Link]

Chen, S.-T., 2013: Multiclass support vector classification to estimate typhoon rainfall distribution. *Disaster Adv.*, **6**, 110-121.

Chen, S.-T., 2015a: Grey rainfall-stage model for forecasting flash flood stages during typhoons. *Disaster Adv.*, **8**, 11-19.

Chen, S.-T., 2015b: Mining informative hydrologic data by using support vector machines and elucidating mined data according to information entropy. *Entropy*, **17**, 1023-1041, doi: 10.3390/e17031023. [Link]

Chen, S.-T., 2019: Probabilistic forecasting of coastal wave height during typhoon warning period using machine learning methods. *J. Hydroinform.*, **21**, 343-358, doi: 10.2166/hydro.2019.115. [Link]

Chen, S.-T. and P.-S. Yu, 2007a: Real-time probabilistic forecasting of flood stages. *J. Hydrol.*, **340**, 63-77, doi: 10.1016/j.jhydrol.2007.04.008. [Link]

Chen, S.-T. and P.-S. Yu, 2007b: Pruning of support vector networks on flood forecasting. *J. Hydrol.*, **347**, 67-78, doi: 10.1016/j.jhydrol.2007.08.029. [Link]

Chen, S.-T., P.-S. Yu, and Y.-H. Tang, 2010: Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *J. Hydrol.*, **385**, 13-22, doi: 10.1016/j.jhydrol.2010.01.021. [Link]

Chen, S.-T., P.-S. Yu, and B.-W. Liu, 2011: Comparison of neural network architectures and inputs for radar rainfall adjustment for typhoon events. *J. Hydrol.*, **405**, 150-160, doi: 10.1016/j.jhydrol.2011.05.017. [Link]

Chen, S.-T., T.-C. Yang, C.-M. Kuo, C.-H. Kuo, and P.-S. Yu, 2013b: Probabilistic drought forecasting in Southern Taiwan using El Niño-Southern Oscillation index. *Terr. Atmos. Ocean. Sci.*, **24**, 911-924, doi: 10.3319/TAO.2013.06.04.01(Hy). [Link]

Dawson, C. W. and R. Wilby, 1998: An artificial neural network approach to rainfall-runoff modelling. *Hydrol. Sci. J.*, **43**, 47-66, doi: 10.1080/02626669809492102. [Link]

Doong, D.-J., S.-T. Chen, Y.-C. Chen, and C.-H. Tsai, 2020: Operational probabilistic forecasting of coastal freak waves by using an artificial neural network. *J. Mar. Sci. Eng.*, **8**, 165, doi: 10.3390/jmse8030165. [Link]

Han, D., L. Chan, and N. Zhu, 2007: Flood forecasting using support vector machines. *J. Hydroinform.*, **9**, 267-276, doi: 10.2166/hydro.2007.027. [Link]

Hsiao, L.-F., D.-S. Chen, J.-S. Hong, T.-C. Yeh, and C.-T. Fong, 2020: Improvement of the Numerical Tropical Cyclone Prediction System at the Central Weather Bureau of Taiwan: TWRF (Typhoon WRF). *Atmosphere*, **11**, 657, doi: 10.3390/atmos11060657. [Link]

Jhong, Y.-D., C.-S. Chen, H.-P. Lin, and S.-T. Chen, 2018: Physical hybrid neural network model to forecast typhoon floods. *Water*, **10**, 632, doi: 10.3390/w10050632. [Link]

Kao, I.-F., Y. Zhou, L.-C. Chang, and F.-J. Chang, 2020: Exploring a Long Short-Term Memory based Encoder-Decoder framework for multi-step-ahead flood forecasting. *J. Hydrol.*, **583**, 124631, doi: 10.1016/j.jhydrol.2020.124631. [Link]

Kao, I.-F., J.-Y. Liou, M.-H. Lee, and F.-J. Chang, 2021: Fusing stacked autoencoder and long short-term memory for regional multistep-ahead flood inundation forecasts. *J. Hydrol.*, **598**, 126371, doi: 10.1016/j.jhydrol.2021.126371. [Link]

Kim, S., J. Jeong, M. Zohaib, and M. Choi, 2018: Spatial disaggregation of ASCAT soil moisture under all sky condition using support vector machine. *Stoch. Environ. Res. Risk Assess.*, **32**, 3455-3473, doi: 10.1007/s00477-018-1620-3. [Link]

Krzysztofowicz, R., 2001: The case for probabilistic forecasting in hydrology. *J. Hydrol.*, **249**, 2-9, doi: 10.1016/S0022-1694(01)00420-6. [Link]

Lin, G.-F. and M.-C. Wu, 2009: A hybrid neural network model for typhoon-rainfall forecasting. *J. Hydrol.*, **375**, 450-458, doi: 10.1016/j.jhydrol.2009.06.047. [Link]

Lin, G.-F. and M.-C. Wu, 2011: An RBF network with a two-step learning algorithm for developing a reservoir inflow forecasting model. *J. Hydrol.*, **405**, 439-450, doi: 10.1016/j.jhydrol.2011.05.042. [Link]

Lin, G.-F., Y.-C. Chou, and M.-C. Wu, 2013a: Typhoon flood forecasting using integrated two-stage support vector machine approach. *J. Hydrol.*, **486**, 334-342, doi: 10.1016/j.jhydrol.2013.02.012. [Link]

Lin, G.-F., B.-C. Jhong, and C.-C. Chang, 2013b: Development of an effective data-driven model for hourly typhoon rainfall forecasting. *J. Hydrol.*, **495**, 52-63, doi: 10.1016/j.jhydrol.2013.04.050. [Link]

Luitel, B., G. Villarini, and G. A. Vecchi, 2018: Verification of the skill of numerical weather prediction models in forecasting rainfall from U.S. landfalling tropical cyclones. *J. Hydrol.*, **556**, 1026-1037, doi: 10.1016/j.jhydrol.2016.09.019. [Link]

Moeeni, H. and H. Bonakdari, 2017: Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model. *Stoch. Environ. Res. Risk Assess.*, **31**, 1997-2010, doi: 10.1007/s00477-016-1273-z. [Link]

Moses, O. and S. Ramotonto, 2018: Assessing forecasting models on prediction of the tropical cyclone Dineo and the associated rainfall over Botswana. *Weather Clim. Extremes*, **21**, 102-109, doi: 10.1016/j.wace.2018.07.004. [Link]

Nguyen, D. T. and S.-T. Chen, 2020: Real-time probabilistic flood forecasting using multiple machine learning methods. *Water*, **12**, 787, doi: 10.3390/w12030787. [Link]

Sattari, M. T., K. Yurekli, and M. Pal, 2012: Performance evaluation of artificial neural network approaches in forecasting reservoir inflow. *Appl. Math. Model.*, **36**, 2649-2657, doi: 10.1016/j.apm.2011.09.048. [Link]

Teschl, R., W. L. Randeu, and F. Teschl, 2007: Improving weather radar estimates of rainfall using feed-forward neural networks. *Neural Netw.*, **20**, 519-527, doi: 10.1016/j.neunet.2007.04.005. [Link]

Villarini, G., W. F. Krajewski, A. A. Ntelekos, K. P. Georgakakos, and J. A. Smith, 2010: Towards probabilistic forecasting of flash floods: The combined effects of uncertainty in radar-rainfall and flash flood guidance. *J. Hydrol.*, **394**, 275-284, doi: 10.1016/j.jhydrol.2010.02.014. [Link]

Wu, M.-C. and G.-F. Lin, 2017: The very short-term rainfall forecasting for a mountainous watershed by means of an ensemble numerical weather prediction system in Taiwan. *J. Hydrol.*, **546**, 60-70, doi: 10.1016/j.jhydrol.2017.01.012. [Link]

Wu, M.-C., S.-C. Yang, T.-H. Yang, and H.-M. Kao, 2018: Typhoon rainfall forecasting by means of ensemble numerical weather predictions with a GA-based integration strategy. *Atmosphere*, **9**, 425, doi: 10.3390/atmos9110425. [Link]

Yang, T.-C., P.-S. Yu, C.-M. Wei, and S.-T. Chen, 2011: Projection of climate change for daily precipitation: A case study in Shih-Men Reservoir catchment in Taiwan. *Hydrol. Process.*, **25**, 1342-1354, doi: 10.1002/

hyp.7883. [Link]

Yu, P.-S. and S.-T. Chen, 2005: Updating real-time flood forecasting using a fuzzy rule-based model. *Hydrol. Sci. J.*, **50**, 265-278, doi: 10.1623/hysj.50.2.265.61796. [Link]

Yu, P.-S., C.-J. Chen, and S.-J. Chen, 2000: Application of gray and fuzzy methods for rainfall forecasting. *J. Hydrol. Eng.*, **5**, 339-345, doi: 10.1061/(ASCE)1084-0699(2000)5:4(339). [Link]

Yu, P.-S., C.-J. Chen, S.-J. Chen, and S.-C. Lin, 2001: Application of Grey model toward runoff forecasting. *J. Am. Water Resour. Assoc.*, **37**, 151-166, doi: 10.1111/j.1752-1688.2001.tb05482.x. [Link]

Yu, P.-S., S.-T. Chen, C.-C. Wu, and S.-C. Lin, 2004: Comparison of grey and phase-space rainfall forecasting

models using a fuzzy decision method. *Hydrol. Sci. J.*, **49**, 655-672, doi: 10.1623/hysj.49.4.655.54427. [Link]

Yu, P.-S., S.-T. Chen, C.-J. Chen, and T.-C. Yang, 2005: The potential of fuzzy multi-objective model for rainfall forecasting from typhoons. *Nat. Hazards*, **34**, 131-150, doi: 10.1007/s11069-004-8889-x. [Link]

Yu, P.-S., S.-T. Chen, and I.-F. Chang, 2006: Support vector regression for real-time flood stage forecasting. *J. Hydrol.*, **328**, 704-716, doi: 10.1016/j.jhydrol.2006.01.021. [Link]

Zhou, Y., S. Guo, C.-Y. Xu, F.-J. Chang, and J. Yin, 2020: Improving the reliability of probabilistic multi-step-ahead flood forecasting by fusing unscented Kalman filter with recurrent neural network. *Water*, **12**, 578, doi: 10.3390/w12020578. [Link]